User selection of optimal HRTF sets via holistic comparative evaluation

Rishi Shukla[1], Rebecca Stewart[1], Agnieszka Roginska[2], and Mark Sandler[1]

[1]*Centre for Digital Music, Queen Mary University of London, London, UK*
[2]*Music and Audio Research Lab, New York University, New York, NY, USA*

This paper was presented at the Audio Engineering Society Conference on Audio for Virtual and Augmented Reality 2018, as paper number P4-2. The full published version can be found at:
(http://www.aes.org/e-lib/inst/browse.cfm?elib=19677)

ABSTRACT

If well-matched to a given listener, head-related transfer functions (HRTFs) that have not been individually measured can still present relatively effective auditory scenes compared to renderings from individualised HRTF sets. We present and assess a system for HRTF selection that relies on holistic judgements of users to identify their optimal match through a series of pairwise adversarial comparisons. The mechanism resulted in clear preference for a single HRTF set in a majority of cases. Where this did not occur, randomised selection between equally judged HRTFs did not significantly impact user performance in a subsequent listening task. This approach is shown to be equally effective for both novice and expert listeners in selecting their preferred HRTF set.

1 Introduction

There are numerous challenges in capturing bespoke HRTF data for the purpose of binaural synthesis, including barriers related to cost, time, expertise and specialised resources. Subjective selection of preferred sets from a database of HRTF measurements is recognised as a plausible alternative means of accommodating individual requirements for spatial audio rendering over headphones [1, 2]. The primary shortcoming of this approach manifests as increased front-back confusion for users (i.e. erroneous perception of sources rendered in the front field as coming from the back) [3, 4]. Incorporation of head-tracking is shown to mitigate this by allowing subtle or subconscious head rotation to verify virtual source positions [5].

Any virtual reality (VR), augmented reality (AR) or mixed reality (MR) system deploying binaural synthesis to a mass user base could therefore benefit from an effective method for users to select the non-individualised HRTF set that works best for them personally. Considerations for the efficacy of any selection system would include:

- *Reliability* – Does the system return an identifiable preference in a significant majority of cases? For a non-critical task undertaken by an end user, we suggest that a 90-95% success rate might be considered desirable.
- *Vaildity* – Does the returned HRTF provide sufficiently faithful spatial rendering for the user?
- *Usability* – Can the system be operated equally successfully by any user, irrespective of listening expertise?
- *Efficiency* – Is the overall time taken to complete the selection process of an acceptable duration? For the purposes of single-time calibration of a recreation-focussed system, we suggest that no more than ten minutes might be considered desirable and under five minutes ideal or preferable.

Traditional subjective HRTF evaluation deploys localisation testing to gain a granular view of spatial distortions that occur when a specific set is used by any one listener. In these cases, participants are played a series of sound sources rendered binaurally and asked to make an absolute judgement on perceived virtual locations. Extents and patterns of localisation error are examined to assess the suitability of the HRTF set. This approach has been used to inform understanding of non-individualised HRTFs' limitations [3, 4, 5] and has also been applied to demonstrate how users can potentially be trained, over time, to learn and interpret more accurately the spatialisation cues of generic HRTF sets [6, 7]. However, this assessment method

is too time-consuming to be applied in the context of HRTF selection in an end user system.

More recently developed approaches have used participants' relative judgements to evaluate the apparent effectiveness of an HRTF under a range of criteria and conditions. These typically use qualitative scales for listeners to assess the perceived clarity of changes in specific parameters (such as externalisation, elevation, front-back discrimination, sense of direction, sense of distance, etc.) and have used continuous [8], fixed-point [9, 10] or binary [2, 11, 12, 13] metrics. A key shortcoming identified in these kind of approaches is that they are more reliably applied with expert or familiar users of binaural audio systems [8, 13].

The system outlined here presents the listener with pairs of HRTF sets and asks them to select a preference in each case. Rather than either absolute or relative parametric judgements, the method uses an interactive, holistic evaluation to determine, for each pair, which works best for the listener. The outcomes of each selection round are then used to sort the collection into a final ranked order.

Utilising comparative judgements has already been established in psychophysical research as an effective means of assigning rank to any stimuli that must be evaluated according to a subjective perceptual response [14]. Pairwise comparison has also been previously used in [15] as a proposed means of selecting non-individualised HRTFs. In their study, each participant started with a collection of 32 HRTF sets selected randomly from a pool of 120. They ran an adapted Swiss-style tournament (where a winner is determined using aggregated points accumulation), which eliminated any twice defeated HRTFs – meaning that not all possible pairings were presented to the listener. The comparison task used a one second pink noise burst stimulus presented in an incremental orbit, at locations 30° apart on the horizontal plane (0° elevation).

In contrast, our approach exhaustively iterates over every possible pairwise HRTF set combination and uses recorded music tracks as stimuli within an interactive system. The next section describes the method in more detail and how it was evaluated with 22 users. The subsequent section presents results of the evaluation and is followed by a discussion of our findings.

2 Methodology

This study evaluates a method for selecting the preferred non-individualized HRTF set for binaural audio synthesis from a collection of HRTF sets. The study is divided into two parts:

1. Participants compare pairs of binaurally synthesised spatial renderings of a single song presented over headphones. For each pair, either render is of the same song but convolved with one of two HRTF sets implementing first order horizontal-only virtual Ambisonics [16]. The participant can choose either HRTF as they are listening at any given time and rotate the song's virtual position around their head using an interactive interface.

2. Participants complete 64 music search or browsing tasks by navigating through a two-dimensional binaural auditory scene containing 15 songs. The scene is generated using the same spatialisation technique as in part one. The binaural signal is rendered using the HRTF set assigned to the participant from the outcome of the selection process. Tasks are presented in a variety of configurations. However, overall outcomes from only the search task trials are described and discussed here as they are the most pertinent to evaluating the HRTF selection procedure in part one.

Head-tracking is not used during any part of the study.

2.1 HRTF Selection Tournament

Six HRTF sets were identified for use in the study – three from the *LISTEN* [17] and three from the *CIPIC* databases [18]. These sets are the three from each collection that performed best under horizontal plane localization tests in [2].

Use of six HRTF sets results in 15 pairwise comparisons. Pairs are determined by a round robin tournament structure, where every one of the six available HRTF sets meets each of its five opponents once using randomly generated scheduling. A round robin tournament allows us to determine, definitively, which HRTF set(s) within the group of six are preferred by a participant. It is purposefully distinct from the approach used in [15], as it establishes a ground truth and data set that

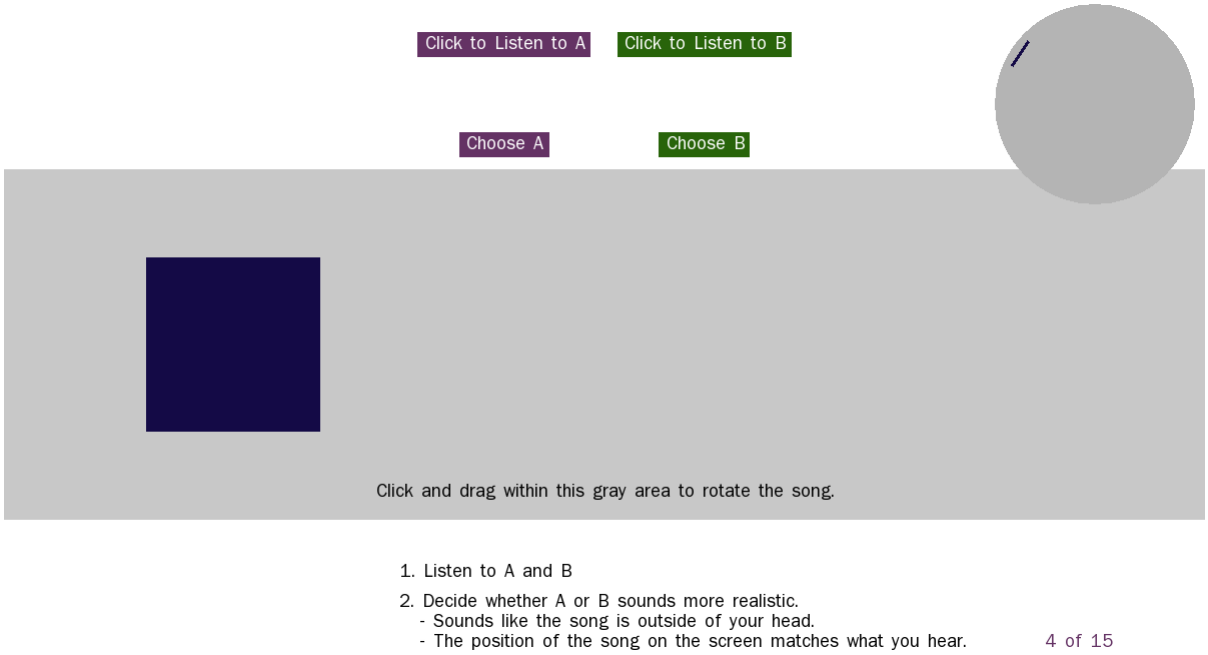

Fig. 1: Graphical user interface for the HRTF selection tournament.

can be used to model and evaluate less comprehensive tournament formats.

For each tournament round, the participant is asked to choose the HRTF set that provides the better spatial effect. To do this, they are instructed to think holistically about the realism of the spatialisation, giving specific attention to sense of externalisation and accuracy of localisation. Figure 1 shows the graphical user interface (GUI) and text prompt used to elicit responses.

Only one song is presented at a time. The participant interacts with the grey area in the GUI using a mouse to rotate the spatial image clockwise (drag right) or anticlockwise (drag left). When the sound source is in the front field (i.e. within +/- 90° azimuth) a blue square is displayed within the grey area relative to its orbital location. The circle in the upper right corner of the window illustrates the active song's spatial position from an overhead perspective, which updates in real-time with user interaction. When shifted beyond +/- 180° azimuth, the active song transitions out of the auditory scene and one of two other songs is rotated into the environment in the concurrent direction. By including this interactive mechanism within the comparison process we aim to simulate features of potential multimedia applications, where user control and spatial scene rotation might (ideally) be integral to the listening and sound localisation experience.

The same three songs are presented for each of the 15 pairings. Each song is of the same musical genre (latin pop), is edited to fade at one minute and plays back on a repeated loop. Before starting the tournament, the participant is shown a video to demonstrate how to interact with the GUI and fully explain the task they need to complete. This includes the instruction:

> "Listen to A and B. Rotate the song and decide if A or B has a better 3D audio effect. Consider whether the song sounds outside your head, if it sounds like it really goes behind and in front of you, and if where you hear the song matches where you see the song on the screen."

Participants are also informed that this is a calibration process and that there is no right or wrong answer, only their own personal preference.

As shown in Figure 1, the navigation and selection software enables A/B comparisons by allowing a seamless

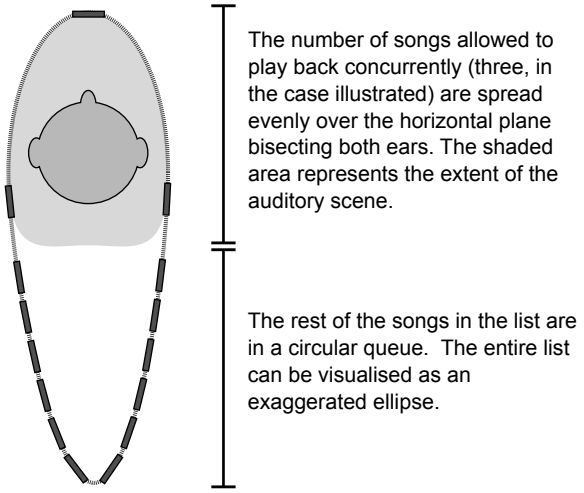

Fig. 2: Search task auditory environment.

switch between HRTF sets without changing the song's virtual position in space or playback point. After 15 rounds the HRTF set chosen most frequently as the winner of a tournament round (up to a maximum of five times) is identified as the participant's preferred set. If there is no clear winner, one HRTF set is randomly selected from the tied top results.

2.2 Auditory Navigation Trials

After choosing their winning HRTF set, the participant is presented with 64 auditory navigation trials using a binaural audio interface rendered using their preferred HRTF set. Each trial is randomly generated as either a search or browse task and both tasks use a modified version of the GUI used from the HRTF comparison process (Figure 2). Only search task data is used in aggregate form for further analysis and discussion here.

The search task requires the participant to correctly identify and select a specific song from within a spatialised auditory scene containing 15 tracks (illustrated in Figure 2). For each search trial, either one, two, three or four tracks are rendered concurrently in positions evenly distributed across the horizontal plane bisecting the participant's ears. The participant rotates manually through the entire queue of 15 sound sources to locate the required track (illustrated in Figure 3). As in the HRTF selection system, when any song is rotated beyond the position immediately behind the listener (i.e. +/- 180° azimuth) it falls out of the auditory scene and is replaced by the next one in the elliptical queue of 15.

Two further conditions are varied for search tasks: graphical visualisations of song locations (on or off) and the genre of songs presented (mixed, rock, hip-hop and jazz). All three variables (number of concurrent songs, graphical visualisation and genre) are allocated randomly. As with the HRTF selection tournament, before the navigation trials start the participant is shown a second video outlining the context of the listening test they are about to undertake and an explanation of its accompanying GUI. For every search task, data for both trial outcome (correct or incorrect selection) and time taken is logged by the testing system.

At this stage we note that a poorly matched HRTF would be expected to make the majority of auditory search tasks more challenging for users. With two, three and four concurrent songs, less accurate binaural rendering would introduce greater front/back confusion and reduce clarity in sound source localisation. This would be compounded in the 50% of tasks where a visual representation of sound source locations is not displayed. Such conditions could be anticipated to either increase selection error (reduce success), or make disorientation more likely (lengthen response time).

2.3 Software Implementation

The software engine – i.e. the audio rendering, tournament scheduler, trial generation engine and logging system – is developed in Python. The GUI is developed using openFrameworks and communicates with the engine via Open Sound Control.

2.4 Participants

Twenty-two volunteers participated in the study and were paid for their time at the standard rate set by the university. Subjects were drawn from a combination of departmental staff and students at the authors' two institutions and from open public calls in New York and London. Eight participants identified as female and 14 as male. Twenty-one of the volunteers' ages were distributed across five brackets ranging from 18-24 to 40-44 and one was aged over 60. All participants self-reported that they did not have a hearing impairment.

The two parts of the experiment were conducted in a dedicated room with quiet surroundings and took around two hours to complete, including a ten minute break. Each subject completed the experiment in one

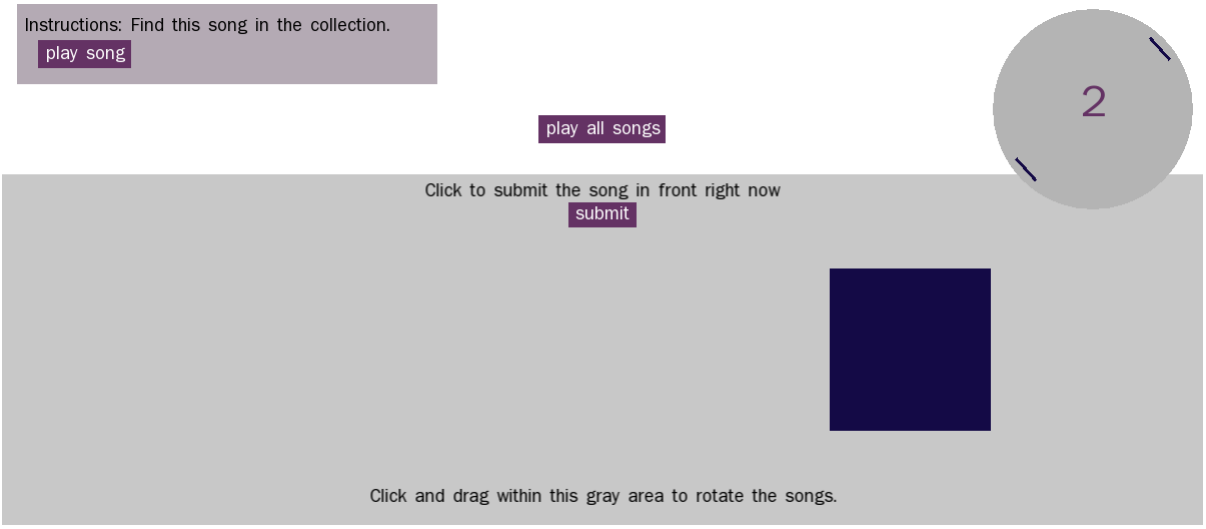

Fig. 3: Graphical user interface for the auditory search task (with visual location information included).

session and also filled in a questionnaire, which requested information about their prior musical experience and music listening habits (reported in section 3.3). From part two of the experiment, the 22 subjects generated 701 search task trial data points, which ranged from 28-34 per participant. This variation is due to the randomised presentation of search or browse type tasks that made up each participant's total of 64.

3 Results

From the first part of the study, we look at how often an HRTF set is chosen in individual tournament rounds, when it was ranked as an overall winner and the strength of each outcome. In combination with data collected from the second part of the study, we examine HRTF selection strength against participants' performance in search tasks. An overview of participant listening habits and musical experience is also presented in relation to HRTF selection strength. We then perform a comparison of the round robin results with the projected outcome of a knock-out tournament.

3.1 HRTF selections

Figure 4 shows that when the individual tournament match outcomes from each participant are viewed collectively, one HRTF set performs clearly below chance level (*CIPIC 58*, with 22.7%) and another notably above (*CIPIC 15*, with 63.6%). One other set performs marginally under chance level (*LISTEN 1014*) and the rest slightly over. A Friedman test confirms there is a significant difference in the mean overall popularity of *CIPIC 58* compared to *LISTEN 1022* ($p = 0.01$), *LISTEN 1028* ($p = 0.007$), *CIPIC 12* ($p = 0.03$) and *CIPIC 15* ($p = 0.002$). *LISTEN 1014*, on the other hand, is not significantly different in its mean ranking to either *CIPIC 58* or any of the top four sets.

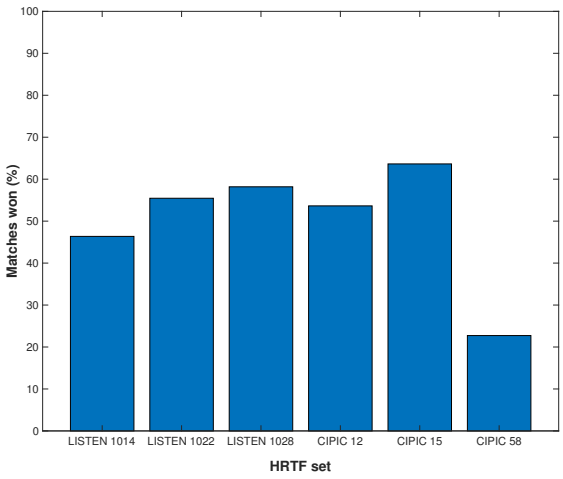

Fig. 4: Percentage of matches won by each HRTF set.

To analyse the degree of certainty or strength of the winning HRTF set two indices were used: winning HRTF set score and winning HRTF set margin. The

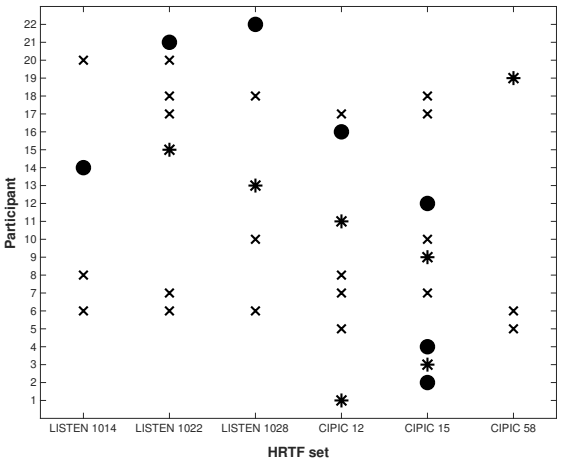

Fig. 5: Breakdown of winning HRTF sets (x = tie, $*$ = winner by margin of 1, $\bullet$ = winner by margin of 2).

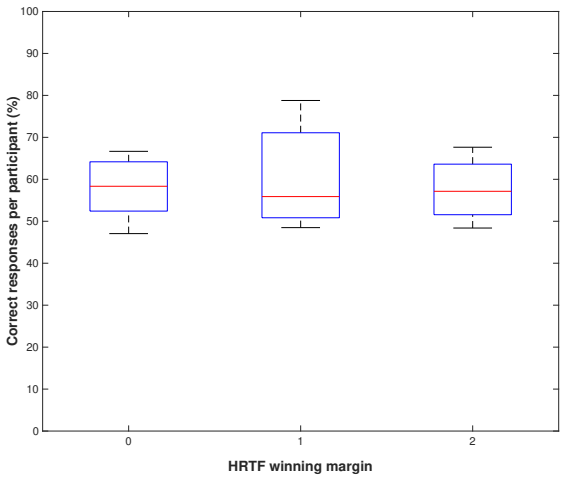

Fig. 6: Correct search task responses per participant, grouped by HRTF winning margin.

winning score is the number of tournament round wins attained by the selected HRTF set, which has a value of either 3/5, 4/5 or 5/5. The winning margin is the difference in tournament round wins between the selected HRTF and the second place set, which has a value of either 0 (in the event of a tie), 1 or 2. The distribution of both measures is shown in Table1. For all but one participant, the competitive selection process resulted in a winning score of either 5/5 or 4/5.

	Margin			***Totals***
	0	**1**	**2**	
Score of 3/5	1	0	0	***1***
Score of 4/5	7	2	0	***9***
Score of 5/5	0	5	7	***12***
Totals	***8***	***7***	***7***	

Table 1: Distribution of HRTF selection results.

The aggregate performance of each HRTF (found in Figure 4) is mirrored in the outcomes for individual participants (in Figure 5). For instance, *CIPIC 15* was the most frequently chosen set in all individual matches and the most commonly chosen as an outright or joint overall winner (five and four times, respectively). Both of the least selected HRTFs overall were still outright winners in one case (*CIPIC 58*, for participant 19 and *LISTEN 1014*, for participant 14). Furthermore, in each of these instances, the HRTF was an undefeated winner, scoring 5/5 in the selection process.

The instructional video told participants that the HRTF selection process would take approximately 10 minutes and in practice the average time taken to complete it was a mean of 11.36 and median of 9.5 minutes.

3.2 HRTF selection strength and user performance

Chi-square tests between the number of correct search task responses per participant and the three winning score groups of the HRTF competition process (3/5, 4/5 or 5/5) show no significant effect (χ^2 = 0.028; p = 0.986). ANOVA tests of all search task response times show no significant difference between the same groups (F = 1.85; p = 0.159). There was therefore no evident relationship between HRTF winning score and task performance.

When comparing winning score to the winning margin as indicators of selection strength, winning score is a less useful metric for two reasons reflected in Table 1. First, membership of the categories is particularly imbalanced at 1, 9 and 12 participants in each group. Second, the 4/5 category contains a number of tied results whereby the HRTF set used was subsequently selected randomly amongst the top tied winners. For these reasons, the winning margin is instead identified as the preferred index for selection strength and is the measure referred to in the rest of this section.

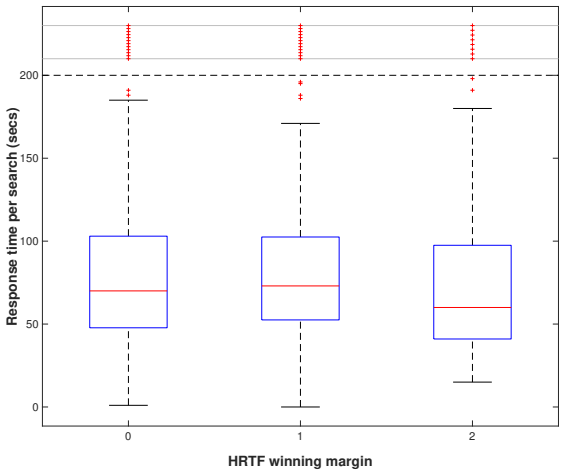

Fig. 7: All search task response times, grouped by HRTF winning margin.

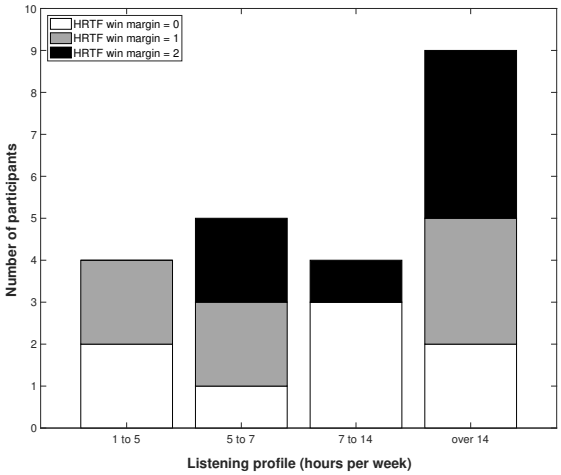

Fig. 8: Makeup of winning margin groups, by participant listening profile.

Chi-square tests of correct search task responses per participant between the three winning margin groups of the HRTF tournament selection (0, 1 or 2) also show no significant effect (χ^2 = 0.688; p = 0.709), which is reflected in Figure 6. ANOVA tests of all search task response times show differences between the same groups only at p<0.1 (F = 2.52; p = 0.081). This slight trend, shown in Figure 7, is in line with expectations – i.e. participants with the strongest HRTF selection outcome tended to respond quicker on average, but not to a significant degree.

3.3 HRTF selection strength and user expertise

The makeup of the selection strength groups was also cross-referenced against personal listening habits and musical experience. The breakdown in Figures 8 and 9 show fairly even representation of all winning margin groups across levels of listening and musical training. Chi-square tests show no significant difference in the makeup of winning margin groups against either factor.

3.4 Alternate tournament format

A round robin tournament is an exhaustive competitive tournament structure, as all possible combinations are evaluated. This is in comparison to knock-out tournaments where only consecutive winners are matched against each other. The latter format results in fewer comparisons and therefore a shorter overall tournament. We simulated a knock-out tournament to compare if this shorter format would result in different winning HRTF sets than the results of the round robin format.

Each participant's tournament was re-run using the same sequence of HRTF pairs and results as in the live experiment, but any losing HRTF set was eliminated from future rounds. Subsequent comparisons that involved an eliminated HRTF set were then ignored. This tournament model reduced the number of comparisons per participant from 15 to just five – with each round removing one of the six HRTFs from the competition.

This re-projected format resulted in a different final selection for just six of the 22 participants. Furthermore, for five of these six, the knock-out winner had also been a joint winner in the round robin (i.e. had won the same number of comparisons) and had not been designated as the chosen HRTF merely due to random selection between top tied results. Thus only one participant would have actually been allocated a weaker choice under this identification method. In that instance, the participant would have been allocated an HRTF that they had deemed favourable in only 2/5 comparisons, rather than the winning score of 4/5 that had resulted from their completed round robin tournament.

4 Discussion

To inform our discussion, we return to the four criteria put forward in section one for evaluating the success of an HRTF selection system.

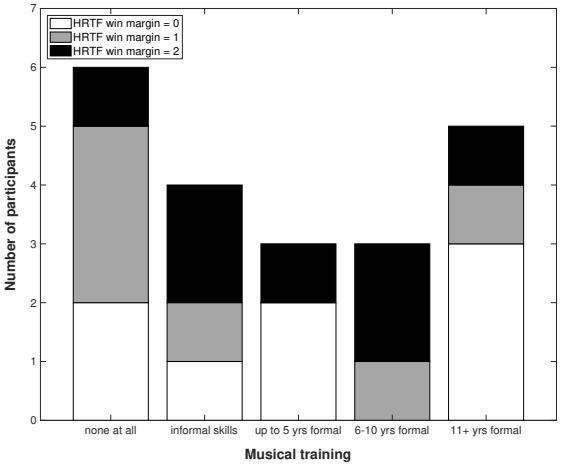

Fig. 9: Makeup of winning margin groups, by participant musical training. (Note that one participant did not report on their musical training.)

4.1 Reliability

Participants frequently showed demonstrable preference for an HRTF set through the comparative selection process, with 95% of tournaments resulting in a score of 4/5 (either as an individual or joint winner) or 5/5. Moreover, the distribution of competition results for HRTFs overall and by individual participants confirms that less preferred sets, in the context of a larger population, can nevertheless be well-matched for specific individuals. This pattern is consistent with previous research [2, 11] and suggests that the requirements of listeners whose best fitting HRTF set is less commonly chosen in aggregate can still be successfully matched under this system.

This data attests to the effectiveness of the pairwise comparison mechanism for consistently identifying personally preferred HRTF sets. Had the system been less effective, we would expect to see a higher proportion of unclear outcomes. There is only one instance of either a joint winning score of 3, or a four-way tie (participant 6, in both cases). There were only three further cases of a three-way tie (participants 7, 17 and 18).

4.2 Validity

The initial findings presented here show that the relative strength of the HRTF match had no significant influence on user performance in the subsequent auditory retrieval trials. This is the case if participant selections are analysed either by winning score, or winning margin and assessed in terms of search task success rates or response times. In particular, the relative difference in response time between those participants allocated a random selection from two or more tied HRTF sets, compared to those who had selected a very clear singular preference, is not significant.

This data provides some verification as to the validity of the proposed HRTF selection process for task-oriented applications, such as in the one devised for this research. I.e. all strengths of selection returned by the system (winning margin of 0, 1 or 2) returned a fit that produced similar levels of accuracy and speed, on average. However, more data analysis and research is required to understand the full relationship between the quality of HRTF fit and its affect on auditory navigation trial performance.

This study also demonstrates a holistic, interactive approach to subjective selection of non-individualised HRTFs. In doing so, it contributes to wider discussions on how quality of binaural rendering systems can be assessed using means that extend beyond measurements of localisation accuracy alone [19, 20, 21].

4.3 Usability

There is no evidence from this participant group to suggest that either personal listening habits or musical experience influenced their ability to make a more or less decisive HRTF selection via the method tested.

4.4 Efficiency

Preliminary investigation indicates that the round robin tournament format could be shortened and still achieve very similar outcomes with potentially only small impact on reliability. Reducing the process by a third (and in theory from around 10-11 minutes to 3-4 minutes in duration) only altered one outcome substantively.

5 Summary

We have outlined and tested a mechanism for the selection of a non-individualized HRTF set based on holistic comparative judgements by users of an interactive binaural system without head-tracking. This system has been shown to result in consistent identification of optimal HRTFs – whether or not a singular preference

is ultimately identified by the process. The outcomes of the selection method have been tested with a task-oriented simulation. Results show no significant effect between different strengths of HRTF preference that resulted and task competence.

Moreover, neither personal listening patterns nor musical training appear to influence the strength of HRTF choice that presents from using this selection mechanism, demonstrating its potential applicability to both novice and expert users. We have also shown how more efficient tournament structures might be used in future systems to gain selections with a similar level of certainty, but more rapidly and thus to potentially incorporate greater choice of HRTF sets.

Further work will be concentrated in two main areas. First, additional analysis is still to be conducted on the full variety of auditory navigation tasks pursued by participants having made their HRTF selection. Detailed examination of this data might provide further evidence regarding the relative performance of participants' HRTF selections. For instance, correlation between HRTF selection strength and success rate or time taken might be more significant if the more challenging task conditions (i.e. multiple songs, no visualisation and/or a single genre) are analysed in isolation.

Second, it is worth restating that the holistic judgement drawn out from participants in this solution only caters for 2D binaural display. More research needs to be dedicated to how holistic evaluation of a 3D binaural effect would be elicited, both in terms of the design of stimulus material and the wording of the pairwise comparison asked of users.

In conclusion, this study supports an interactive and iterative calibration process that allows users to choose an optimal non-individualised HRTF set. When looking towards end-user applications, the process needs to: result in clearly identifiable preference(s) for the desired proportion of cases (be reliable); return an HRTF set that provides a faithful spatial image (be valid); be equally effective for expert and non-expert listeners (be usable); not burden the user through excessive and lengthy tests (be efficient). The method presented here has been shown to fulfil all of these aims.

6 Acknowledgements

This work was supported by EPSRC and AHRC under the grant EP/L01632X/1 (Centre for Doctoral Training in Media and Arts Technology) and by ESPRC Platform Grant EP/E045235/1.

References

[1] Seeber, B. U. and Fastl, H., "Subjective selection of non-individual head-related transfer functions," in *Proceedings of the 2003 International Conference on Auditory Display*, pp. 259–262, ICAD, Boston, MA, USA, 2003.

[2] Roginska, A., Wakefield, G. H., and Santoro, T. S., "User selected HRTFs: Reduced complexity and improved perception," in *Undersea Human System Integration Symposium*, pp. 1–14, Providence, RI, USA, 2010.

[3] Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L., "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, 94(1), pp. 111–123, 1993.

[4] Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D., "Binaural technique: Do we need individual recordings?" *Journal of the Audio Engineering Society*, 44(6), pp. 451–469, 1996.

[5] Begault, D. R., Wenzel, E. M., and Anderson, M. R., "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source." *Journal of the Audio Engineering Society*, 49(10), pp. 904–916, 2001.

[6] Medonca, C., Santos, J. A., Campos, G., Dias, P., Vieira, J., and Ferreira, J., "On the improvement of auditory accuracy with non-individualized HRTF-based sounds," in *Audio Engineering Society Convention 129*, pp. 1–8, San Francisco, CA, USA, 2010.

[7] Berger, C. C., Gonzalez-Franco, M., Tajadura-Jiménez, A., Florencio, D., and Zhang, Z., "Generic HRTFs may be good enough in virtual reality. Improving source localization through cross-modal plasticity," *Frontiers in Neuroscience*, 12(February), pp. 1–9, 2018.

[8] Schönstein, D. and Katz, B. F. G., "Variability in perceptual evaluation of HRTFs," *Journal of the Audio Engineering Society*, 60(10), pp. 783–793, 2012.

[9] Andreopoulou, A. and Katz, B. F. G., "Investigation on subjective HRTF rating repeatability," in *Audio Engineering Society Convention 140*, pp. 1–10, Paris, France, 2016.

[10] Katz, B. F. G. and Parseihian, G., "Perceptually based head-related transfer function database optimization," *The Journal of the Acoustical Society of America*, 131(2), pp. 99–105, 2012.

[11] Roginska, A., Santoro, T. S., and Wakefield, G. H., "Stimulus-dependent HRTF preference," in *Audio Engineering Society Convention 129*, pp. 1–11, San Francisco, CA, USA, 2010.

[12] Wan, Y., Zare, A., and Mcmullen, K., "Evaluating the consistency of subjectively selected head-related transfer functions (HRTFs) over time," in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, pp. 1–8, Helsinki, Finland, 2014.

[13] Andreopoulou, A. and Roginska, A., "Evaluating HRTF similarity through subjective assessments : Factors that can affect judgment," *Joint 40th International Computer Music Conference (ICMC) & 11th Sound and Music Computing Conference*, (September), pp. 1375–1381, 2014.

[14] Thurstone, L. L., "A law of comparative judgment," *Psychological Review*, 34(4), pp. 273–286, 1927.

[15] Iwaya, Y., "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," *Acoustical Science and Technology*, 27(6), pp. 340–343, 2006.

[16] Noisternig, M., Musil, T., Sontacchi, A., and Höldrich, R., "3D binaural sound reproduction using a virtual ambisonic approach," in *VECIMS 2003 - 2003 International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems*, pp. 174–178, Lugano, Switzerland, 2003.

[17] Warusfel, O., "Listen HRTF database," `http://recherche.ircam.fr/equipes/salles/listen/index.html`, 2003.

[18] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C., "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*, pp. 99–102, New Paltz, NY, USA, 2001.

[19] Nicol, R., Gros, L., Colomes, C., Warusfel, O., Noisternig, M., Bahu, H., Katz, B. F. G., and Simon, L. S. R., "A roadmap for assessing the quality of experience of 3D audio binaural rendering," in *EAA Joint Symposium on Auralization and Ambisonics*, pp. 100–106, Universitätsverlag der TU Berlin, Berlin, Germany, 2014.

[20] Simon, L. S. R., Zacharov, N., and Katz, B. F. G., "Perceptual attributes for the comparison of head-related transfer functions," *The Journal of the Acoustical Society of America*, 140(5), pp. 3623–3632, 2016.

[21] Reardon, G., Roginska, A., Flanagan, P., Calle, J. S., Genovese, A., Zalles, G., Olko, M., and Jerez, C., "Evaluation of binaural renderers: a methodology," in *Audio Engineering Society Convention 143*, pp. 1–6, New York, NY, USA, 2017.